DOCUMENT RESUME

ED 201 666 TM 810 259

AUTHOR Plake, Barbara S.; And Others
TITLE Number Right Scores as a Function of Item Arrangement, Knowledge of Arrangement, and Test Anxiety.
PUB DATE Apr 81
NOTE 29p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Los Angeles, CA, April 11-17, 1981).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Academic Achievement; Achievement Tests; Difficulty Level; Higher Education; *Mathematics; *Sex Differences; *Test Anxiety; *Test Format; Test Items; Test Wiseness
IDENTIFIERS ACT Assessment; Alpert Haber Achievement Anxiety Test; *Item Position (Tests); Mathematics Anxiety Rating Scale; State Trait Anxiety Inventory (Spielberger)

ABSTRACT

Effects of item arrangement (easy-hard, uniform, and random), test anxiety, and sex on a 48-item multiple-choice mathematics test assembled from items of the American College Testing Program and taken by motivated upper level undergraduates and beginning graduate students were investigated. Four measures of anxiety were used: the Achievement Test Anxiety Scale, STAI-STATE and STAI-TRAIT of Speilberger's State Trait Anxiety Inventory and the Mathematics Anxiety Rating Scale. Significant (P=less than .05) sex, anxiety, and sex by order effects were found for the cognitive variable with males outperforming females, low anxious students scoring higher than anxious ones, and males with easy-hard ordering scoring substantially better than females with the same order of items. Significant anxiety, sex, and order effects were found for the perception scores of perceived difficulty and performance. The fact that males outperformed females on a mathematics test with items arranged easy-hard suggests that item arrangement may be another possible explanation for observed sex differences in mathematics achievement at the college level. (Author/RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED201666

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Number Right Scores as a Function of Item Arrangement, Knowledge of Arrangement, and Test Anxiety

Barbara S. Plake, Stephen R. Lowry, Charles J. Ansorge, and Claire S. Parker

University of Nebraska-Lincoln

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

B. S. Plake

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Running Head: Number Right

Paper presented at the annual meeting of the National Council on Measurement in Education, Los Angeles, 1981

TM 810 219

Abstract

Effects of item arrangement (easy-hard, uniform, and random), test anxiety, and sex on a mathematics test by motivated upper level undergraduates and beginning graduate students were investigated. Significant ($p < .05$) sex, anxiety, and sex by order effects were found for the cognitive variable with males outperforming females, low anxious students scoring higher than anxious ones, and males with easy-hard ordering scoring substantially better than females with that order of items. Significant anxiety, sex, and order effects were found for the perception scores of perceived difficulty and performance. The fact that males outperformed females on a mathematics test with items arranged easy-hard suggests that item arrangement may be another possible explanation for observed sex differences in mathematics achievement at the college level.

Number Right Score as a Function of Item Arrangement, Knowledge of Arrangement, and Test Anxiety

The evaluation of student achievement and aptitude by means of paper and pencil multiple-choice tests is confounded by the presence of test and student characteristics which may obscure the accurate assessment of student performance (Millman, Bishop, & Ebel , 1965). Among test characteristics that may affect test performance are the presentation of the individual questions to the students and the specificity of the test directions. Item arrangement, or the physical location of the test questions within the test is one aspect of item presentation that has been subjected to empirical investigation. Basically, the following four item arrangements have been studied: easy-hard, hard-easy, uniform or spiral-cyclical, and random. Results have generally supported easy-hard ordering for speeded tests, but have shown no practical difference for power tests (Flaugher, Melton, & Meyer, 1968; Brenner, 1964). Specificity of test directions, in particular information provided to the examinee about the item ordering, has been hypothesized to affect test performance (Cronbach, 1946; Hambleton & Traub, 1974), however little empirical evidence has been found to support this idea (Plake, 1980; Plake, Thompson, & Lowry, in press).

Student characteristics which may affect test performance include test wiseness and test anxiety. Studies combining test anxiety and

item arrangement show conflicting results (Hambleton & Traub, 1974; Marso, 1970, Smouse & Munz, 1968, Plake, Thompson, & Lowry, in press). Sarnacki (1979), in his review of factors affecting test performance in the cognitive domain, suggests that familiarity with the arrangement of the items in a test may reduce one's test anxiety. Plake, Thompson, and Lowry (in press) investigated the effects of item arrangement, (easy-hard, uniform, random), knowledge of arrangement and test anxiety on performance on a 48-item mathematics test. No significant ($p > .05$) order or knowledge effects were found. However, the subject pool consisted of college sophomore volunteers from a general educational psychology class who may not have been motivated to achieve on the mathematics test. Therefore, the validity and applicability of the results may be questionable. Sex of the subject has also been suggested as a variable which will influence test performance, particularly in the area of mathematics (Tobias & Weissbrod, 1980).

The purposes of this study were to (1) investigate the effects of item arrangement, knowledge of arrangement, and test anxiety using motivated subjects to contrast with the Plake, Thompson, and Lowry (in press) research results on non-motivated subjects, and (2) investigate the interactive effects of sex on the subject and test anxiety, item arrangement and/or knowledge of arrangement on test performance.

Method

Instruments. A 48-item multiple-choice mathematics test was assembled from items of the ACT: College Mathematics Placement Program to represent a range of items to cover content from elementary mathematics through college algebra. The item difficulty indices ranged from .15 to .84 with a median value of 0.55. The item discrimination indices (point biserial) ranged from .15 to .82 with a median of 0.53.

Three test forms were developed from the 48-item multiple choice test: easy-hard, uniform or spiral cyclical, and random. In the directions which were developed to accompany each test booklet, half contained additional information explaining the order of the items in the form. In the other half of the test booklets, the directions omitted reference to the arrangement of the items in the form. Therefore, six experimental forms were developed: three item arrangements x two conditions of knowledge. These were the same arrangements, directions and test questions used by Plake, Thompson, and Lowry (in press).

Four measures of anxiety were used in this study: Alpert and Haber's (1960) 19-item Achievement Test Anxiety Scale, AAT, Spielberger's STATE and TRAIT Anxiety Inventories, STAI-STATE, STAI-TRAIT, (Spielberger, Gorsuch, & Lushene, 1970) and a 24-item revised Mathematics Anxiety Rating Scale, MARS (Suinn, Edie, Nicolletti, & Spinelli, 1972). The Revised Mathematics Anxiety Rating Scale is a 24-item subset of the original

98-item inventory which has coefficient reliability of .98 and which correlates .97 with total score of the 98 item MARS (Plake & Parker, 1981).

Procedure The six experimental units were administered to 170 students enrolled in three introductory statistics classes at the University of Nebraska-Lincoln. Two were educational psychology classes and one was an agriculture class. The class population consisted of upper level undergraduates and beginning graduate students. The mathematics background and age of the students varied substantially. The students participated in the study as a part of their statistics class requirement. Students were informed that the examination was to be used to identify which students might benefit from a mathematics remediation lab which was scheduled for the following week and available to all students who took the examination. The test materials were administered to each class separately within the time periods assigned to the classes (75 minutes). The instruments were assembled into packets and distributed in a random fashion to the students as they arrived for class. Each packet contained the four anxiety measures (AAT, Revised MARS, STAI-TRAIT, STAI-STATE) arranged in balanced order, followed by the experimental form of the mathematics test. Additional questions were added to the end of each test form to assess the students' perceptions of the test's difficulty and their expected performance. Immediately after completing the experimental form, all students were readministered the STAI-STATE anxiety test.

Analysis. The first analysis was a three factor fixed effects ANOVA with number right as the dependent variable, item order (easy-hard, uniform or spiral cyclical, and random), knowledge of arrangement (informed, uninformed), and sex (male, female) as independent variables and anxiety scores as multiple covariates (AAT-Facilitating, AAT-Debilitating, STAI-TRAIT, pre STAI-STATE, revised MARS).

A second analysis was computed on the same data, this time defining anxiety condition, based on the combination of AAT-Facilitating and Debilitating subscales, in a manner described by Towle and Merrill (1975). This procedure yields four anxiety conditions groups: High facilitating/low debilitating (HL), low facilitating/ high debilitating (LH), high facilitating/high debilitating (HH), and low facilitating/low debilitating (LL). Item order, knowledge of arrangement, sex and anxiety condition were the four independent variables in the ANOVA analysis.

A third analysis focused on the two perception scores (perceived performance and perceived difficulty) in a multivariate analysis of variance. For this analysis item order, sex, and math anxiety condition (lo, hi based on a median split of the revised MARS distribution) were the independent variables.

A four factor repeated measures ANOVA (order, knowledge of arrangement, sex and time) was then used to investigate if any of the experimental forms significantly altered scores from the initial

administrations of the STATE scale (time 1) prior to the mathematics test to the second administration (time 2) afterwards.

An analysis of the omit pattern was achieved by calculating the number of omits per item for the three experimental forms. A χ^2 test for relationship was used to investigate an effect of the item ordering on the omit behavior of the examinees.

In addition, reliability indices (KR_{20}) for the six experimental units were computed and compared for population equality.

Results

The means and standard deviations for the number right (NR), perceived difficulty (PD), and perceived performance (PP) scores for combinations of item order and knowledge of arrangement are shown in Table 1.

The results of the three factor fixed effects ANOVA (the independent variables were item order, knowledge of ordering, sex; the dependent variable was number right on the mathematic test; multiple covariates were AAT-Facilitating, AAT-Debilitating, STAI-TRAIT, pre STAT-STATE, revised MARS) can be found in Table 2.

Insert Tables 1 and 2 about here

Significant main effects for sex ($F(1,153) = 9.86$, $p < .002$), and a sex by order interaction ($F(2,153) = 5.12$, $p < .007$) were found. Males outperformed females on the mathematics test ($\bar{X}_M = 27.60$, $\bar{X}_F = 21.53$). Figure 1 displays the cell means contributing to the sex by order interaction. Simple effect followups for the significant sex by order interaction revealed that, for the easy-hard ordering of the mathematics items, males performed significantly better than females ($X_{E-H,M} = 33.50$,

$\bar{X}_{E-H,F}$ = 20.56; t_{153} = 9.04, p<.001). Males also scored significantly better than the females in the random item ordering ($\bar{X}_{R,M}$ = 25.11, $\bar{X}_{R,F}$ = 20.32, t_{153} = 3.26, p<.01) but no significant simple effect for sex was found for the uniform or spiral cyclical ordering ($\bar{X}_{U,M}$ = 25.75, $\bar{X}_{U,F}$ = 23.62, t_{153} = 1.08, p > .15).

The analysis of NR scores, using item order, knowledge of orderings, sex, and anxiety condition, provided results consistent with those reported above. Significant main effects for sex and anxiety condition were found in addition to a significant sex by order interaction. Results can be found in Table 3. For the four anxiety conditions, followup Tukey pairwise mean comparisons showed a significant difference (α = .05) in performance between the HH group and LL group, ($\bar{X}_{HH}$ = 27.5, $\bar{X}_{LL}$ = 19.32). No other significant pairwise comparisons were found. No significant higher order interactions between anxiety condition, sex and/or ordering of items or knowledge thereof were found.

Insert Table 3 about here

Since the two dependent perception scores correlated .46 with each other, a multivariate ANOVA was selected to analyze the effects of item order, sex and mathematics anxiety on perceived performance and perceived difficulty of the test. Knowledge of arrangement was dropped

from this analysis for two reasons. First, no significant effects for knowledge of arrangement had been shown in any of the analysis of the cognitive variable number right. Second, empty cells had been found in the design when knowledge of ordering was included. This problem may have occurred due to the location of the two perception items within the test booklets. The items were added to the last booklets on a separate sheet of paper and numbered 49 and 50. Some students failed to respond to these items possibly because they knew, from the directions, that there were 48 mathematics items and therefore did not turn to the last page of the test booklet after they had responded to question number 48. In addition, approximately 20% of the students were unable, in the time allotted, to complete the mathematics test and therefore never read the perception questions. Also, some students were confused as to where to mark their responses to these questions and, therefore, failed to code their answers to the perception questions as questions 49 and 50 on the mathematics test answer sheet. The sample size for the subjects who did respond to items 49 and 50 on their mathematics test answer sheet was 114 (reduced from the total of 170 subjects). Due to the location of the questions at the end of the mathematics test, one might speculate that this subset of students would probably be the most mathematically able in the group since they

must have completed the mathematics test in order to address the perception items. A follow-up analysis comparing the test results of the 114 student subgroup who responded to the perception scores to the total group, however, did not confirm the speculation. Under none of the combinations of item arrangement and knowledge of arrangement condition did the mean scores for the 114 student subgroup differ significantly ($\alpha = .05$) from the total group. These results are shown in Table 4. The mathematics test had been used in a similar testing situation with sophomore volunteers, who for the most part, were able to complete the examination in the allotted time.

Insert Table 4 about here

The multivariate analysis found significant main effects for sex (Rao's $F(2,101) = 15.38$, $p < .0001$), order (Rao's $F(4,202) = 3.62$, $p < .007$), and mathematics anxiety (Rao's $F(2,101) = 7.70$, $p < .0008$). The multivariate test for sex by order interaction was not significant at the $\alpha = .05$ level, but would have been at $\alpha = .06$ ($F(4,202) = 2.39$, (Rao's $F(4,202 = 2.39$ $p < .052$). A summary of the MANOVA results are shown in Table 5. Since both variables are of interest, univariate followups, as opposed to stepdowns, were considered. Perceived performance and difficulty both showed significant ($\alpha = .05$) sex effects (Performance: $\bar{X}_M = 3.22$, $\bar{X}_F = 2.0$; Difficulty: $\bar{X}_M = 2.78$, $\bar{X}_F = 2.25$). Males, therefore, viewed the test, regardless of item arrangement, as slightly easier than most tests and expected to perform well on their mathe-

matics test. Females, on the other hand, viewed the test as being more difficult than their male counterparts and were less satisfied with their performance on their mathematics test. Significant order effects ($\alpha = .05$) were found only for perceived difficulty, although the probability value for perceived performance was close to significance ($p < .053$). Due to unequal cell samples sizes, a Dunn followup was used to identify which pairwise item order comparisons were contributing to the significant item order main effect. None of the pairwise comparisons was significant at the $\alpha = .05$ level of significance, ($\bar{X}_{EH} = 2.61$, $\bar{X}_{U} = 2.54$, $\bar{X}_{R} = 2.0$), although by inspection of the means it appears that the major differences in perception are between the easy-hard order and random and the uniform and random, and not between the easy-hard and uniform. Further, the magnitude of the means suggests that, for easy-hard and uniform orderings, students have a slightly lower perception of the difficulty of the test than students with randomly ordered items. Perceived performance and perceived difficulty both showed significant effects ($\alpha = .05$) due to mathematic anxiety. Students with low mathematics anxiety scores felt they had done slightly less than satisfactory and believed their tests were only a little more difficult than most test they had taken before. Students whose test anxiety was high, on the other hand, believed they had performed poorly on the test and reported that the test was

more difficult than most tests they had taken previously.

Insert Table 5 about here

Change in state anxiety, from before the exposure to the experimental unit to after, was analyzed using a four factor ANOVA. Sex, item arrangement, and knowledge of arrangement were fixed effects while time (pre, post) was a within effect. No significant effects were found. These results are summarized in Table 6.

Insert Table 6 about here

Because the physical location of the 48-items varied for the three item arrangements, the locations of items whose pilot test p-values indicated they were difficult were not the same across the test booklets. One tactic that students use when presented with a difficult item is to omit a response. In a test in which items are arranged from easy to hard based on pilot test information, students progress systematically to more and more difficult questions. If, within the same environment, the items are arranged in a different fashion (for example, uniform or random) students encounter the most difficult items throughout the test. If omitting behavior is a function of the difficulty of the item only, and not its relative location within the test, then a consistent

pattern of omitting behavior should be present for the items regardless of the ordering arrangement. In order to determine this, the items were rearranged to match the arrangement of the easy to hard test. The number of omits per item was analyzed using a χ^2 test for relationship. A significant relationship was found ($\chi^2_{94} = 566.14$, $p < .01$) suggesting that the omit behavior of the items is dependent, in part, upon the physical location of the item within the test.

KR_{20} reliabilities were calculated for each of the six experimental forms. These values appear in Table 7. The internal consistency values were tested for population equality using a procedure described by Feldt (1969). None of the reliabilities are significantly different from the others ($\alpha = .05$).

Insert Table 7 about here

Discussion

Item arrangement appears to be an important variable which can, in fact, influence test performance. For mathematics achievement, sex of the subject interacts with item arrangement, yielding highest scores for males with an easy-hard ordering. Perception scores (difficulty and performance) are also influenced by item arrangement. Subjects whose items were arranged easy-hard or uniform had more positive perceptions than subjects whose items were presented in a

random order. Knowledge of arrangement did not appear to significantly influence test performance or perceptions.

Because the content area of the test used in this investigation was mathematics, one might have expected there to be a significant sex by anxiety condition interaction. College aged women have been reported to experience higher levels of mathematics anxiety than similar aged men (Tobias and Weissbrod, 1980; Betz, 1978). A separate analysis, creating high and low mathematics groups based on a median split on the Revised Mathematics Anxiety Rating Scale, was completed on mathematics test performance, using order, knowledge of arrangement, and sex as additional independent variables. Although there were significant sex, anxiety, and sex by order effects, no interactions involving sex and anxiety were found to be significant.

The results of this study are particuarly noteworthy due to the possibility that the test was slightly speeded for the students. It is well established that, when a test is speeded, the easy to hard ordering of the items is best from a psychometric perspective (Mollenkoph, 1950). For this mathematics test with easy to hard item ordering and highly motivated students, males scored significantly better than females, more so than under any of the other item arrangements. Mathematics achievement and aptitude at the national level are often measured by tests which contain items arranged from easy to hard. Male superiority on these tests have been frequently documented

(Fennama & Sherman, 1974; Benbow & Stanley, 1980). A recent study by Benbow and Stanley (1980), using mathematically precocious youth, reported a large sex difference favoring males on the SAT-Math, which was a speeded exam containing items arranged from easy to hard. Perhaps one component of the observed difference in mathematics aptitude between males and females is an artifact of the speeded test conditions and easy-hard item arrangement. It would be interesting to examine whether the combined effects of speed, item arrangement and sex are present in other content areas as well and for other age groups.

Plake, Thompson, and Lowry (in press) presented finding from a study using nonmotivated sophomore subjects which investigated the effects of anxiety, item ordering and knowledge of ordering on mathematics achievement test performance and test perceptions. No significant item arrangement or knowledge of arrangement effects were found. Sex was not a variable in their study. In the present study, using exactly the same mathematics instruments and measures of anxiety, but with motivated upper level students under slightly speeded conditions, and also including sex as a variable in the analysis, significant order by sex effects were found. Whether these results differ from those reported earlier due to the use of motivated students or due to the slight speededness of the testing situation is not clear. Further research, using motivated students under power conditions will help clarify whether

the effects are related to motivation or speed. In addition, the use of a test that measures achievement in a nonmathematical area would help clarify whether the effects are content-area specific. Item arrangement may be identified as an additional source of explanation for the sex differences observed in mathematics achievement.

References

Alpert, R.; & Haber, R.N. Anxiety in Academic Achievement Situation. Journal of Abnormal and Social Psychology, 1960, 61, 207-215.

Betz, N.E. Prevalance, distribution, and correlates of math anxiety in college student. Journal of Counseling Psychology, 1978, 25, 441-448.

Benbow, C.P., & Stanley, J.C. Sex differences in mathematical ability: Factor Artifact. Science, 1980, 210(4475), 1262-1263.

Brenner, M.H. Test Difficulty, Reliability, and Discrimination as Functions of Item Difficulty Order. Journal of Applied Psychology, 1964, 48(2), 98-100.

Cronbach, L.J.. Response Set and Test Validity. Educational and Psychological Measurement, 1964, 6, 475-494.

Feldt, L.S. A Test of the Hypothesis That Cronbach's Alpha or Kuder-Richardson Coefficient Twenty is the Same For Two Tests. Psychometrika, 1969, 34, 363-373.

Fennamma, E. H., & Sherman, J.A. Sex-Related Differences in Mathematics Achievement and Related Factors: A Further Study. Journal for Research in Mathematics Education, May 1978, 189-203.

Flaugher, R.L., & Melton, R.S., & Meyer, C.T. Item Arrangement Under Typical Test Conditions. Education and Psychological Measurement, 1968, 28, 813-824.

Hambleton, R.H., & Traub, R.E. The Effects of Item Order on Test Performance and Stress. Journal of Experimental Education, 1974, 43(1), 40-46.

Millman, J., Bishop, C.H., & Ebel, R. An Analysis of Test Wiseness. Educational and Psychological Measurement, 1965, 25, 707-726.

Mollenkopf, W.G. An Experimental Study of the Effects on Item-Analysis Data of Changing Item Placement and Test Time Limit. Psychometrika, 1950, 15, 291-315.

Plake, B.S. Item arrangement and knowledge of arrangement on test score. Journal of Experimental Education, 1980, 49(1), 56-58.

Plake, B.S., & Parker, C.S. A revised shortened version of the Mathematics Anxiety Rating Scale, 1981, manuscript in preparation.

Plake, B.S., Thompson, P.A., & Lowry, S. Effect of item arrangement, knowledge of arrangement, and test anxiety on two scoring methods. Journal of Experimental Education, in press.

Sarnacki, R.S. An examination of test-wiseness in the cognitive domain. Review of Educational Research, 1979, 49(2), 252-279.

Smouse, A.D., & Munz, D.C. Interaction effects of item-difficulty sequence and achievement-anxiety reaction on anademic performance. Journal of Educational Psychology, 1968, 50(5), 370-374.

Spielberger, C.D., Gorsuch, R.L., & Lushene, R.E. Manual for the State-Trait Anxiety Inventory, Palo Alto: Counsulting Psychologist Press, 1970.

Suinn, R.M., Edie, C.A., Nicoletti, J., & Spinelli, P.R. The MARS, a Measure of Mathematics Anxiety: Psychometrika Data. Journal of Clinical Psychology, 1972, 28, 373-375.

Tobias, S., & Weissbrod, C. Anxiety and Mathematics: An Update. Harvard Educational Review, 1980, 50(1), 63-70.

Towle, N.J., & Merrill, P.F. Effects of anxiety type and item-difficulty sequencing on mathematics test performance. Journal of Educational Measurement, 1975, 4(12), 241-250.

Table 1

Means and Standard Deviations

Number Right (NR), Perceived Difficulty (PD), Perceived Performance (PP)

Informed of Arrangement		Order: E-H $\bar{X}$	E-H sd	E-H n	Uniform $\bar{X}$	Uniform sd	Uniform n	Random $\bar{X}$	Random sd	Random n
	NR	24.89	11.03	28	25.90	10.16	29	23.37	9.21	27
Yes	PD	2.82	0.85	22	2.59	1.00	22	1.75	0.96	12
	PP	2.68	1.49	22	2.41	1.05	22	1.92	1.24	12
No	NR	23.30	9.21	27	22.52	8.81	29	21.73	9.68	30
	PD	2.35	1.11	17	2.50	0.91	22	2.10	1.07	19
	PP	2.47	1.42	17	2.32	1.25	22	2.37	1.42	19

Table 2

ANOVA Summary Table

Number Right (NR) by Item Arrangement, Knowledge of Arrangement, and Sex, With AAT-Facilitating, AAT-Debilitating, STAI-TRAIT, Pre STAT-STATE and Revised MARS and Multiple Covariates

Source	SS	DF	MS	F
Covariates:				
Facilitating	154.16	1	154.16	2.20
Debilitating	122.71	1	122.71	1.75
STAI-TRAIT	20.60	1	20.60	0.29
Pre STAT-STATE	30.85	1	30.85	0.51
Revised MARS	1317.72	1	1317.72	18.81***
Main Effects:				
Item Order	250.35	2	125.18	1.79
Knowledge of Order	120.57	1	120.57	1.72
Sex	690.80	1	690.80	9.86**
Interactions:				
Order x Knowledge	52.08	2	26.04	0.37
Order x Sex	717.24	2	358.62	5.12**
Knowledge x Sex	4.88	1	4.88	0.70
Order x Knowledge x Sex	89.84	2	44.92	0.64
Error	10720.55	153	70.07	

***$p < .001$; **$p < .01$

Table 3

ANOVA Summary Table

Number Right by Item Arrangement, Knowledge of Arrangement, Sex, and Anxiety Condition

Source	SS	DF	MS	F
Item Order	319.82	2	159.91	2.16
Knowledge of Order	176.74	1	176.74	2.38
Sex	1257.92	1	1257.92	16.97***
Anxiety Condition	1179.41	3	393.14	5.30**
Order x Knowledge	41.68	2	20.84	0.28
Order x Sex	813.98	2	406.99	5.49**
Order x Anxiety Condition	649.01	6	108.17	1.46
Knowledge x Sex	62.10	1	62.10	0.84
Knowledge x Anxiety Condition	102.14	3	34.05	0.46
Sex x Anxiety Condition	252.20	3	84.07	1.13
Order x Knowledge x sex	268.07	2	134.03	1.81
Order x Knowledge x Anxiety Condition	549.72	6	91.62	1.24
Order x Sex x Anxiety Condition	559.41	6	93.62	1.26
Knowledge x Sex x Anxiety Condition	166.63	3	55.54	0.75
Order x Knowledge x Sex x Anxiety Condition	194.58	4	48.65	0.66
Error	9191.44	124	74.12	

***p <.001; **p <.01

Table 4

Comparison of Mean Performance on Mathematics Test

For the Subset of Students Responding to the Perception Items

and the Total Subjects by Item Arrangement and Knowledge of Arrangement

Informed of Arrangement		Item Arrangement: E-H		Uniform		Random	
		n	$\bar{X}$	n	$\bar{X}$	n	$\bar{X}$
Yes	Total Group	28	24.89	29	25.90	27	23.37
	Sub-Group	22	26.64	22	28.64	12	23.83
No	Total Group	27	23.30	29	22.52	30	21.73
	Sub-Group	17	23.04	22	23.59	19	25.10

Table 5

MANOVA Summary Table

Perceived Performance, Perceived Difficulty

by Item Arrangement, Sex, and Mathematics Anxiety

Source	Multivariate F's	Univariate F's PP	PD
Order	$F_{4,202}$ = 3.62**	3.02	7.16**
Sex	$F_{2,101}$ = 15.30***	30.86***	9.04**
Anxiety	$F_{2,101}$ = 7.70***	13.13***	9.44**
Order x Sex	$F_{4,202}$ = 2.39	0.80	4.52*
Order x Anxiety Condition	$F_{4,202}$ = 1.01	1.98	0.18
Sex x Anxiety Condition	$F_{2,102}$ = 0.24	0.42	0.30
Order x Sex x Anxiety Condition	$F_{4,202}$ = 1.34	1.65	0.33

***p <.001; **p <.01; *p <.05

Table 6

ANOVA Summary Table

State Anxiety (pre/post) by Item Arrangement

Knowledge of Arrangement and Sex

Source	SS	df	MS	F
Between				
Order	2111.96	2	1055.98	2.78
Knowledge	341.26	1	341.26	0.90
Sex	2.74	1	2.74	0.01
Order x Knowledge	1143.36	2	571.68	1.50
Order x Sex	729.89	2	364.95	0.96
Knowledge x Sex	130.73	1	130.73	0.34
Order x Knowledge x Sex	330.96	2	165.48	0.44
Error (b)	56665.56	149	380.31	
Within				
Time	49.63	1	49.63	0.11
Time x Order	1245.44	2	622.72	1.37
Time x Knowledge	71.38	1	71.38	0.16
Time x Sex	1075.13	1	1075.13	2.36
Time x Order x Knowledge	87.00	2	43.50	0.10
Time x Order x Sex	23.19	2	11.59	0.03
Time x Knowledge x Sex	92.94	1	92.94	0.20
Time x Order x Knowledge x Sex	331.75	2	165.88	0.36
Error (w)	67763.31	149	454.79	

Table 7

Comparison of KR_{20} Reliabilities for Experimental Forms

W+ Values with Reliabilities on Diagonal

		Informed			Not Informed		
		Easy-Hard	Uniform	Random	Easy-Hard	Uniform	Random
	Easy-Hard	0.93					
Informed	Uniform	1.14	0.92				
	Random	1.57	1.37	0.89			
	Easy-Hard	1.43	1.25	1.10	0.90		
Not Informed	Uniform	1.71	1.50	1.09	1.20	0.88	
	Random	1.43	1.25	1.10	1.00	1.20	0.90

$W+ = (1-KR_{20(1)})/(1-KR_{20(2)})$ is approximately distributed as F with df = (N_1-1), (N_2-1)

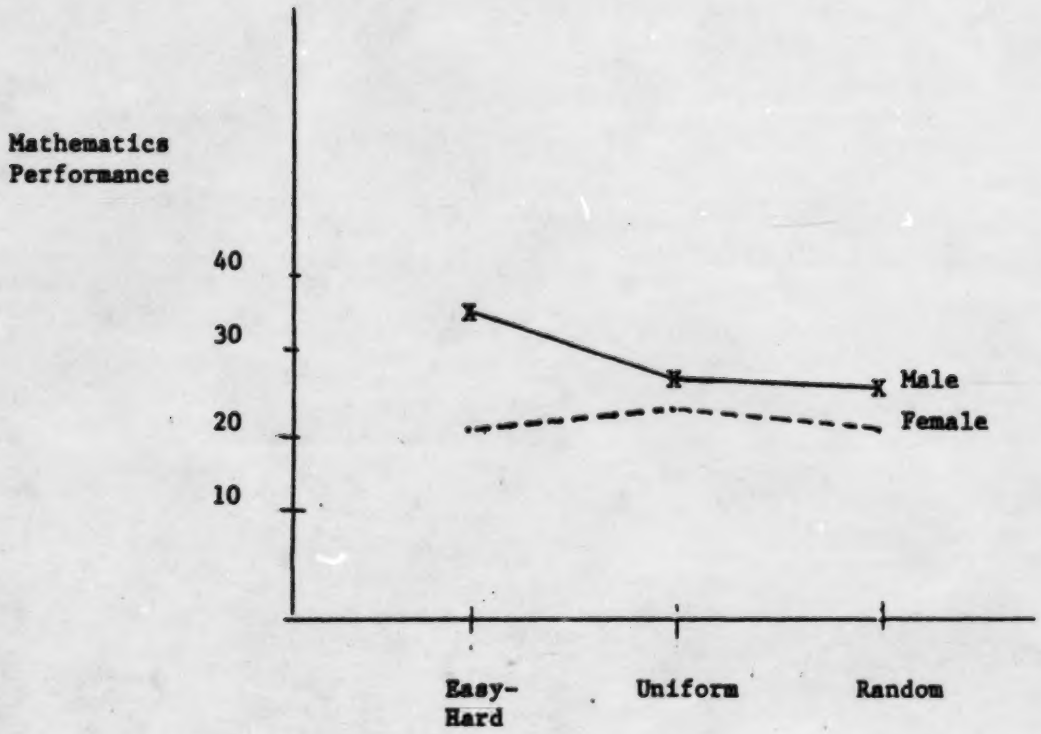

Figure 1. Cell means for sex by order interaction. NR by item arrangement, knowledge of arrangement, and sex analysis.